

Escherichia coli Model System for Studying Gene Expression & Protein Misfolding

Dr. Abdelaziz D. Bior

(Cheyney University of Pennsylvania)

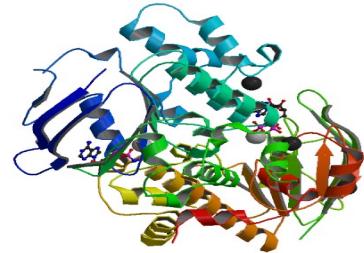
Dr. Paul Freimuth,

(Laboratory , Brookhaven National Laboratory, Department of
Biosciences)

Highlighted Topics of the Talk

- The BNL Synchrotron (NSLS-II) and utilization
- Background
- The Invented cloning technique at BNL
- Hypothesis
- Differences in gene expression and protein synthesis between Prokaryotic and Eukaryotic cells
- Designing of primers and PCR amplification of specific gene sequences
- The issue of alternative reading frames and gain of toxicity
- Summary & Conclusions

Solving Protein Crystal Structures Using NSLS-2

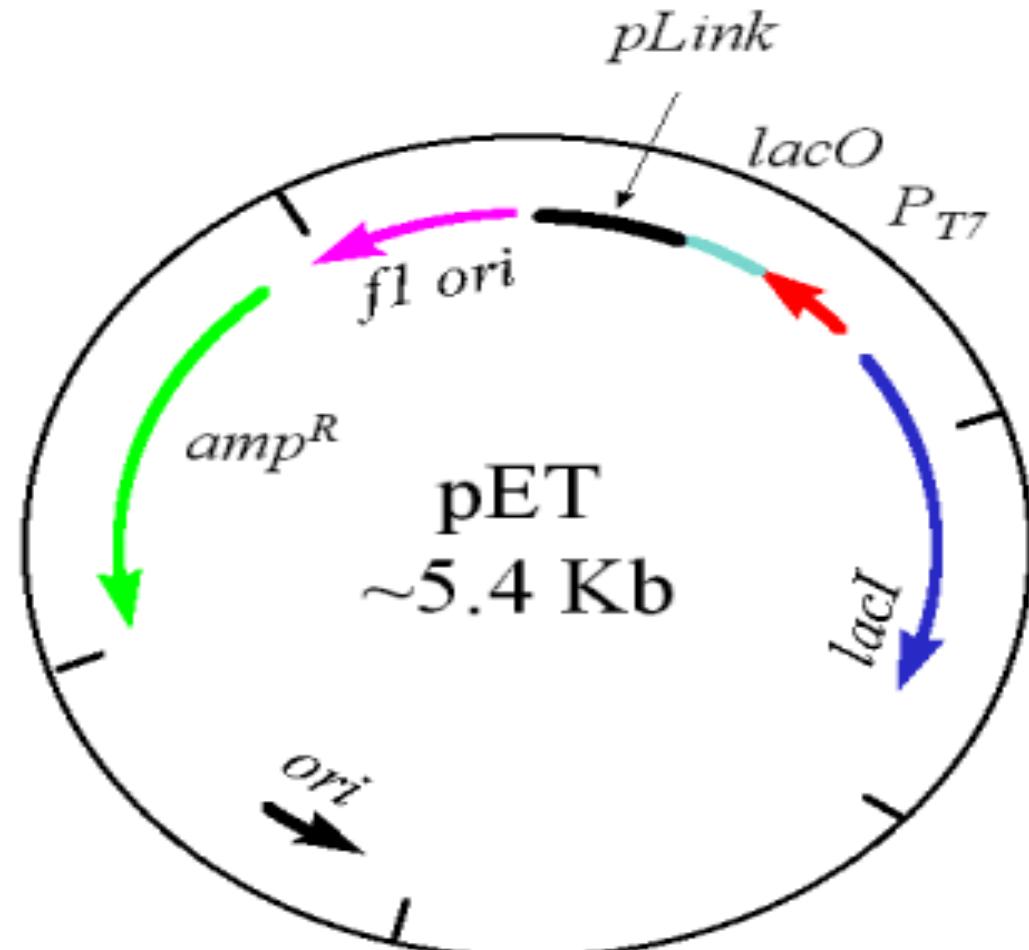


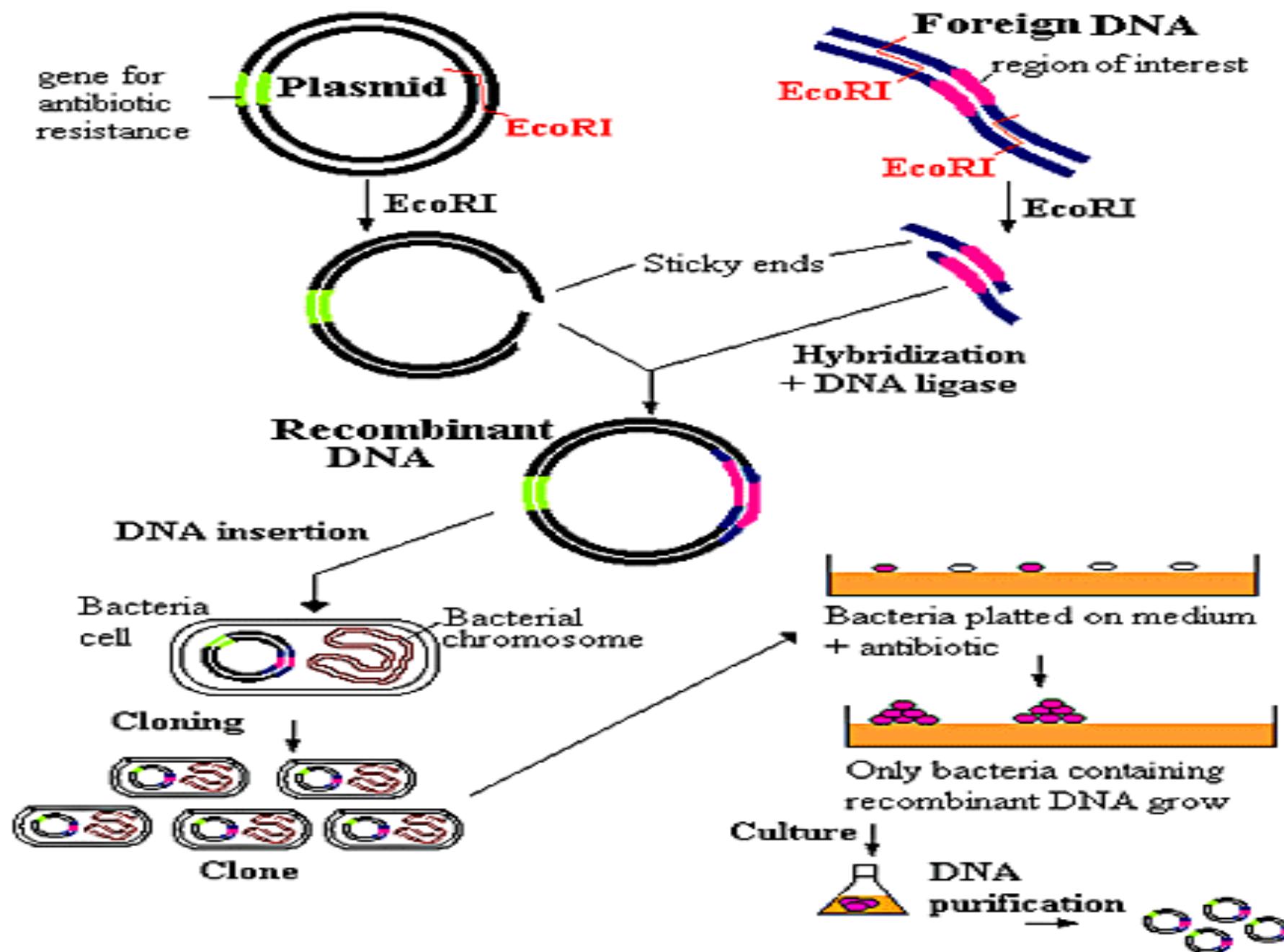
Background

- Expression of cloned genes in prokaryotic and eukaryotic host cells achieves dual purpose:
 - Study of gene function
 - Production of substantial amounts of products for commercial & investigational use.
- Scientific efforts are geared towards establishing strategies to determine appropriate vector-host combination for expression of an exogenous gene
- This is based on many factors that relate ultimately to nature, source and the properties of the gene and its product.
- Currently, there is a problem of compatibility of vector-host combination in reaching the maximum amount of expression of a particular gene
- Proper translation of mRNA transcript is critical to ensure proper folding of the product, and to sustain the protein in the intact and functional state.

The Vector & Host Choice: *Eschrichia coli* Model Expression System Developed in 1986 by W. F. Studier and B. A. Moffatt at BNL

- They created an RNA polymerase expression system which was highly selective for bacteriophage T7 RNA polymerase.
- It was designed to produce many copies of a desired protein within a host cell.
- Recombinant expression vector is inserted (transformed) into a host cell.
- This vector contains all of the genetic coding necessary to produce the protein including the following:
 - A promoter appropriate to the host cell,
 - A sequence which terminates transcription
 - A sequence which codes for ribosome binding (Purves et al., 2001).

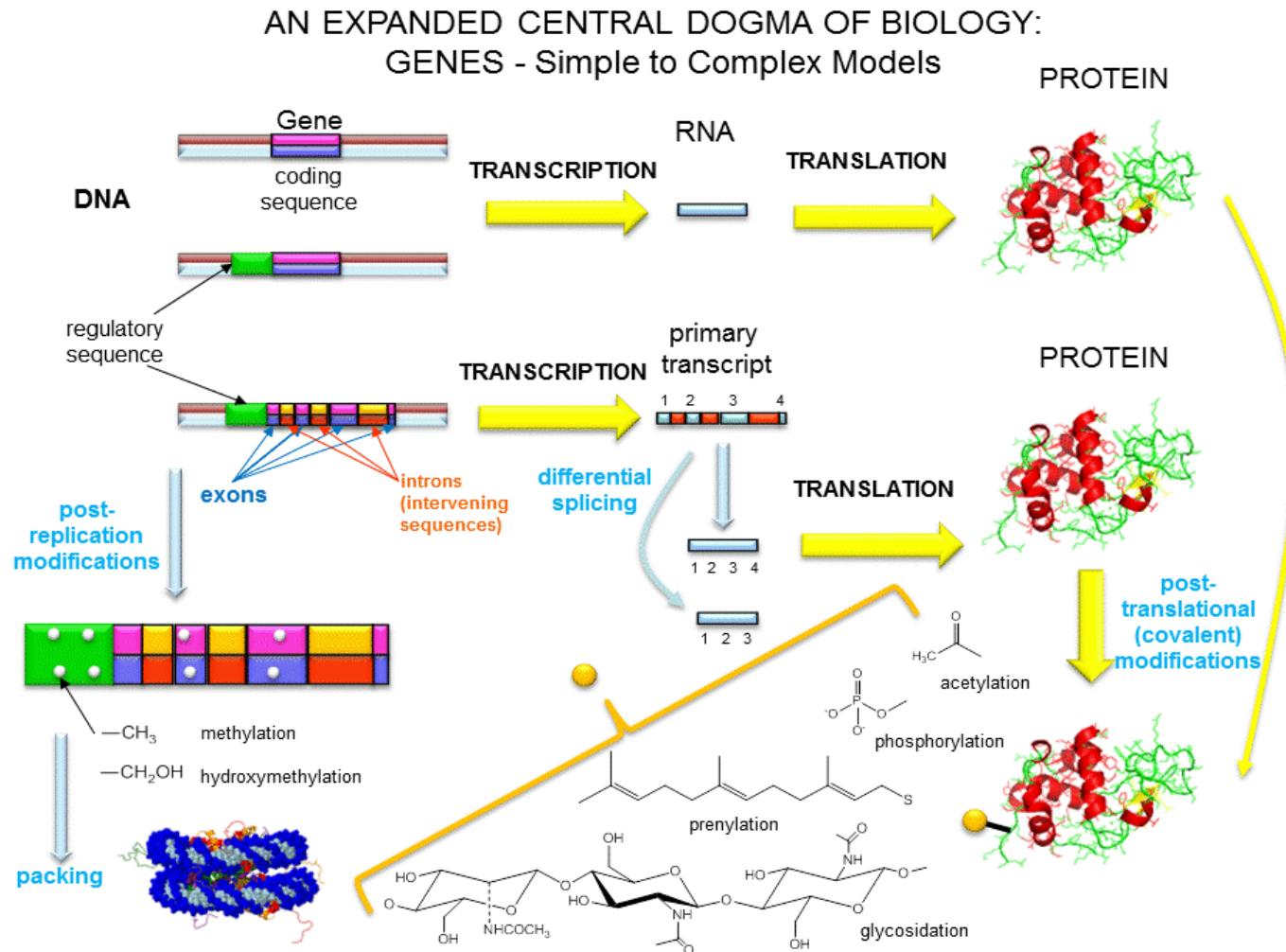
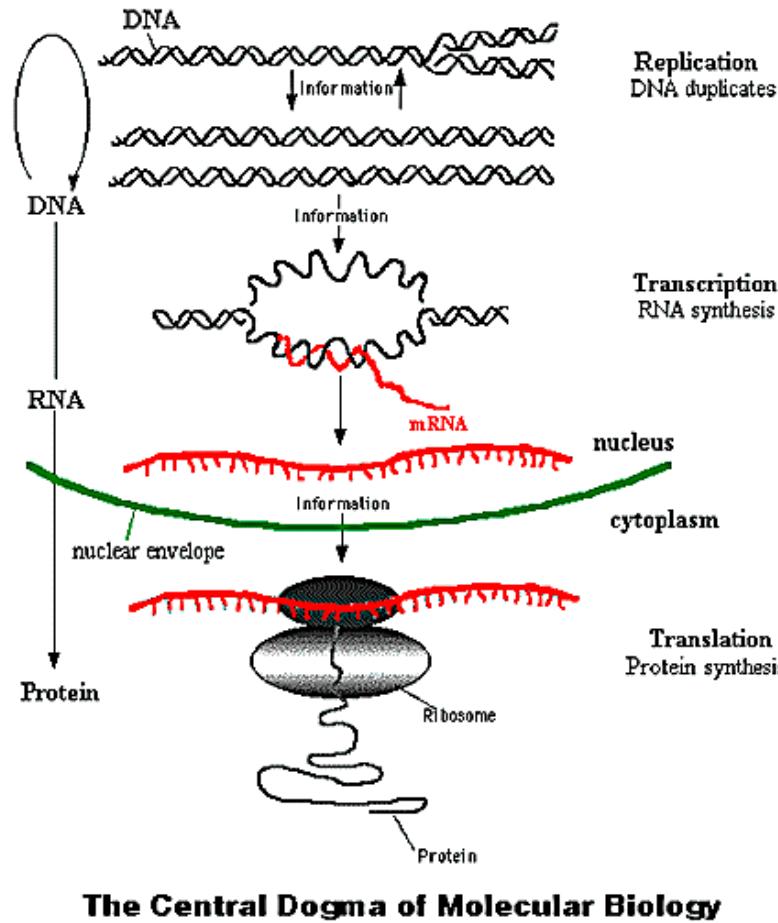




Hypothesis of Summer Project

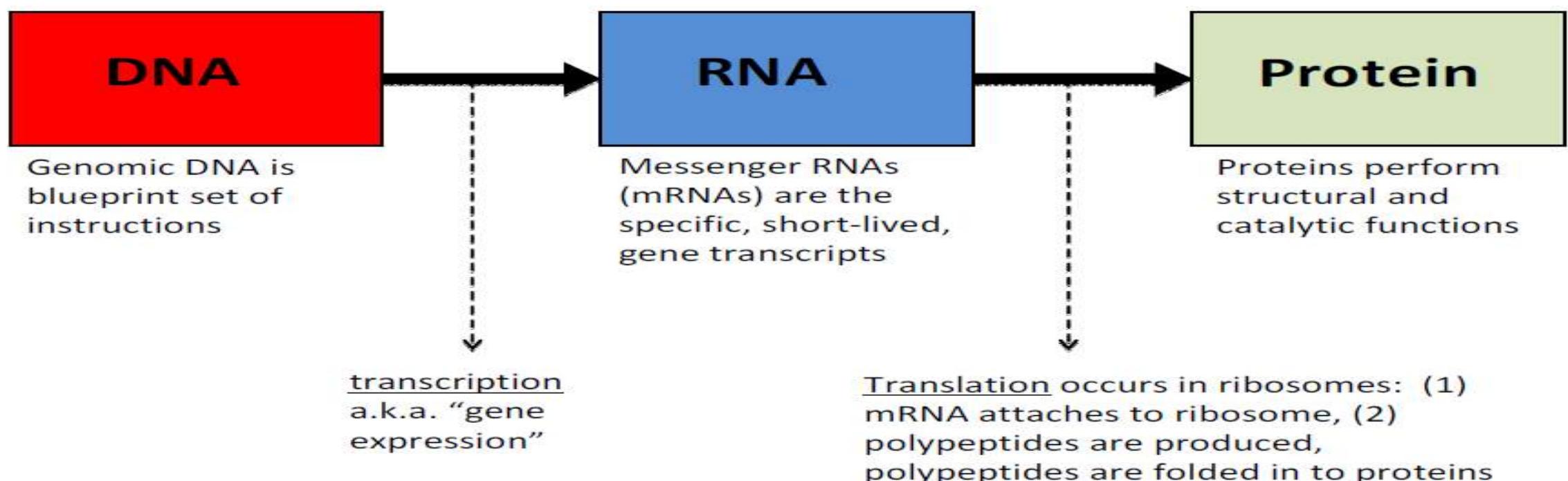
- Expression of enough proteins using the Bacterial expression system developed by Studier at BNL to feed the pipeline for their crystal structure by NSLS-II is still a problem. These proteins are either not made at all, or if made were highly toxic to the cell. We hypothesize that the presence of Shine-Dalgarno Sequences (SD) within the most of eukaryotic genes creates an internal binding sites for ribosomes and initiation of translation downstream of the transcript in bacteria that may produce a toxic peptide to the cells expressed from alternative reading frames. Removal of those sequences upstream of the transcript might enhance production of large amount of nontoxic peptides.

The Central Dogma of Molecular Biology (in Eukaryotic cells)

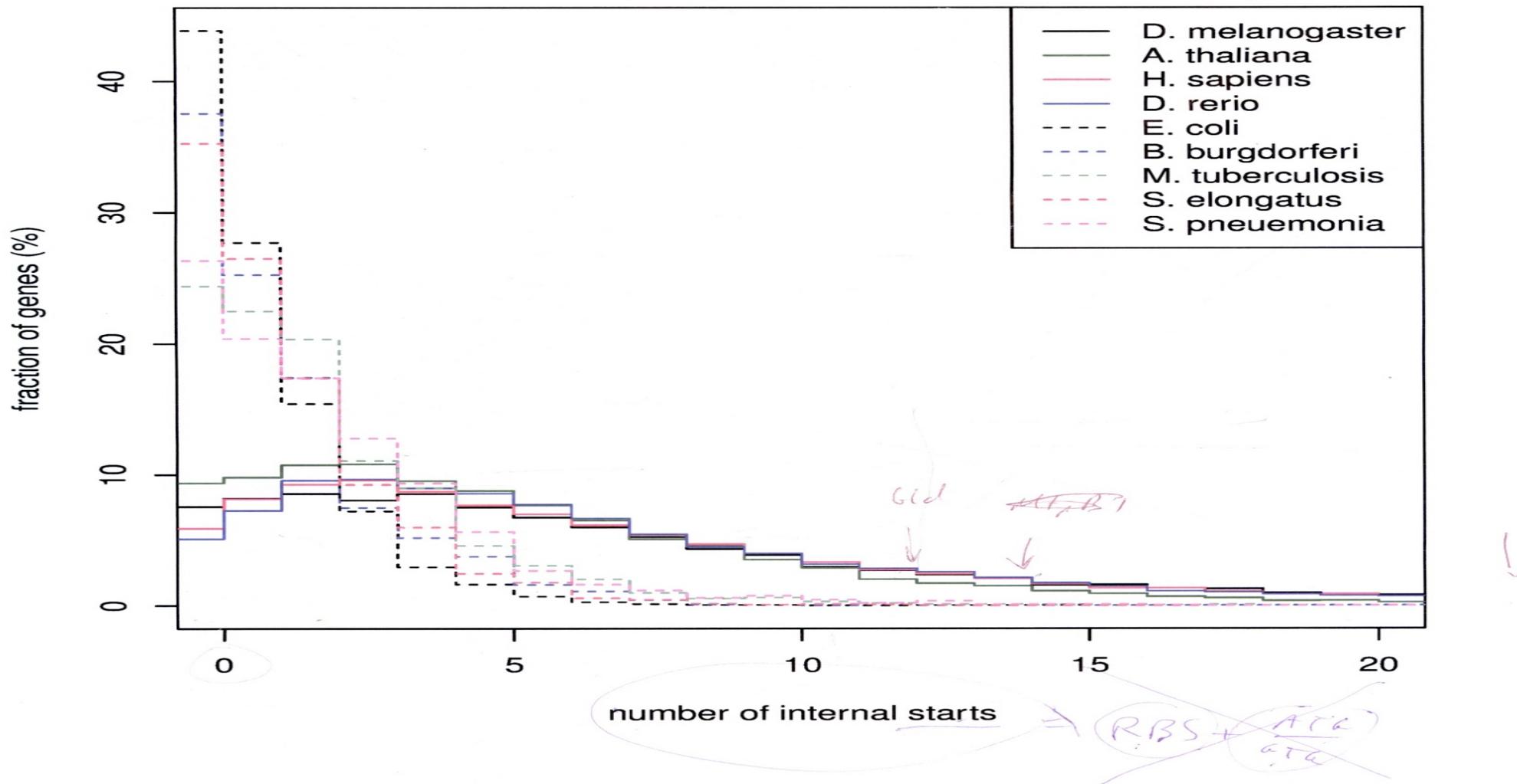


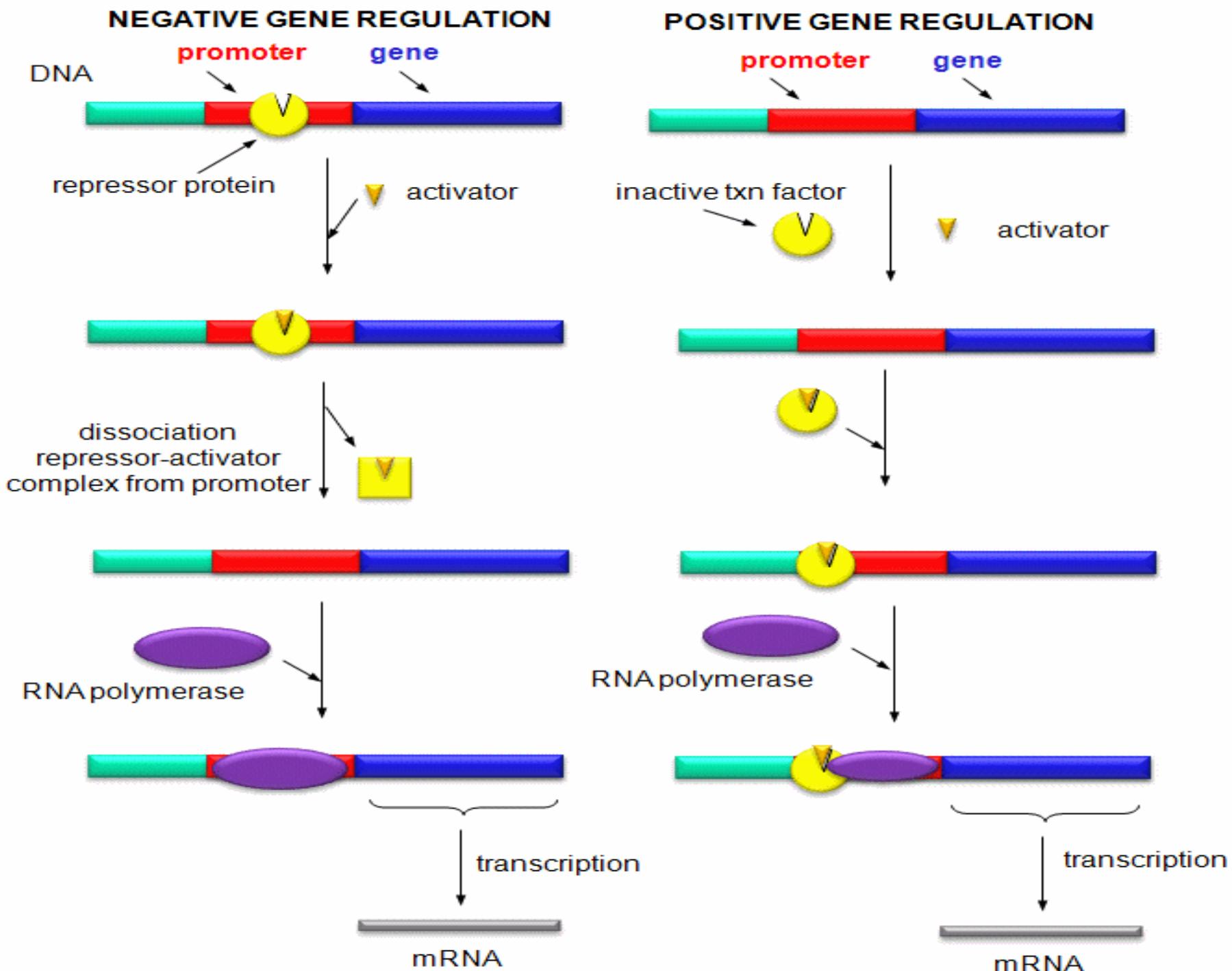
The Central Dogma of Molecular Biology (in Prokaryotic cells)

Central Dogma of Biology:



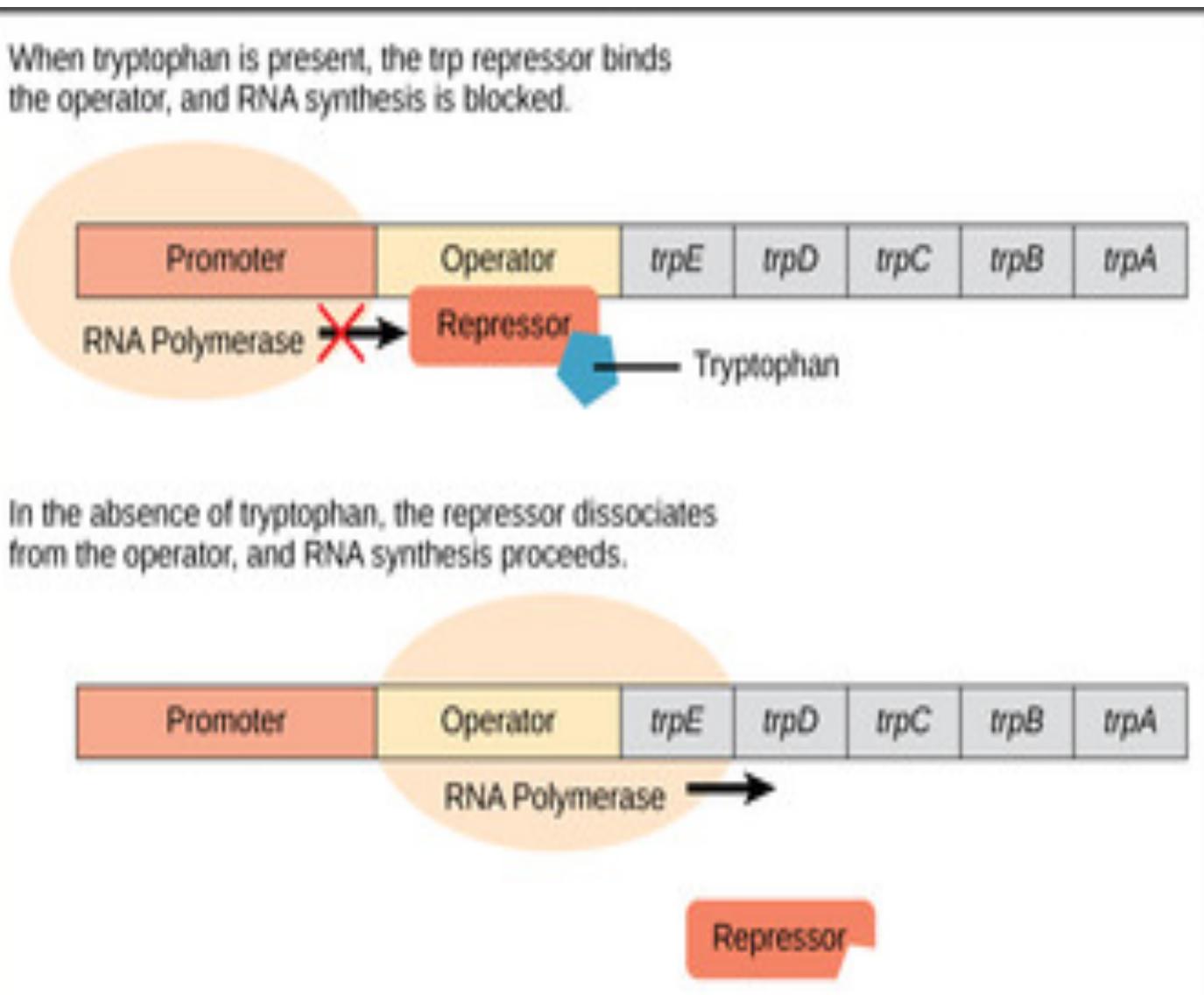
Internal Translation Sites in Prokaryotic & Eukaryotic Genes





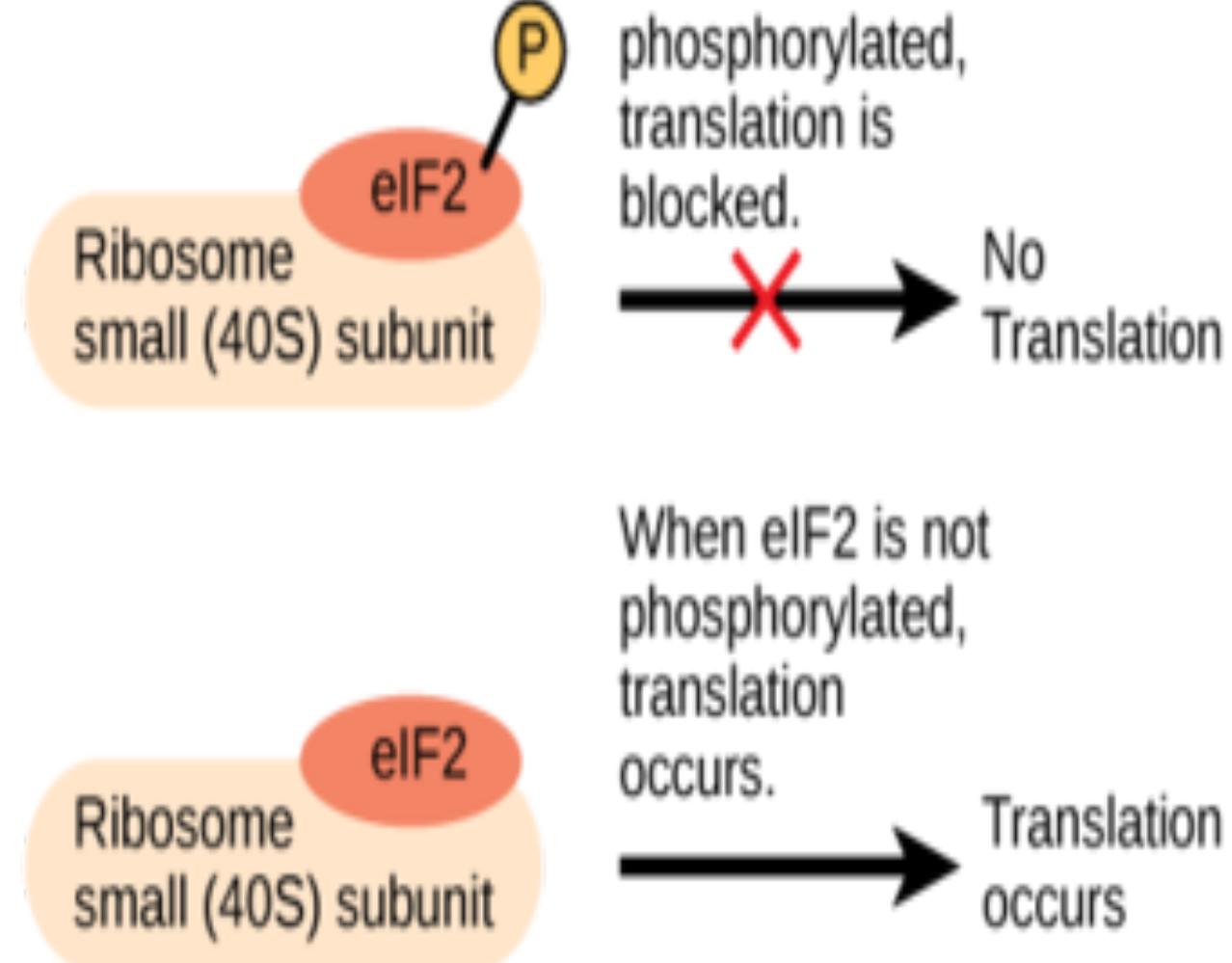
Prokaryotic Gene Expression: The trp Operon

- A Repressor Operon
- Catabolite Activator Protein (CAP): An Activator Regulator
- The lac Operon: An Inducer Operon
- The five genes that are needed to synthesize tryptophan in *E. coli* are located next to each other in the trp-operon.
- When tryptophan is plentiful, two tryptophan molecules bind the repressor protein at the operator sequence.
- This physically blocks the RNA polymerase from transcribing the tryptophan genes.
- When tryptophan is absent, the repressor protein does not bind to the operator and the genes are transcribed.
- Prokaryotes regulate gene expression by controlling the amount of transcription.

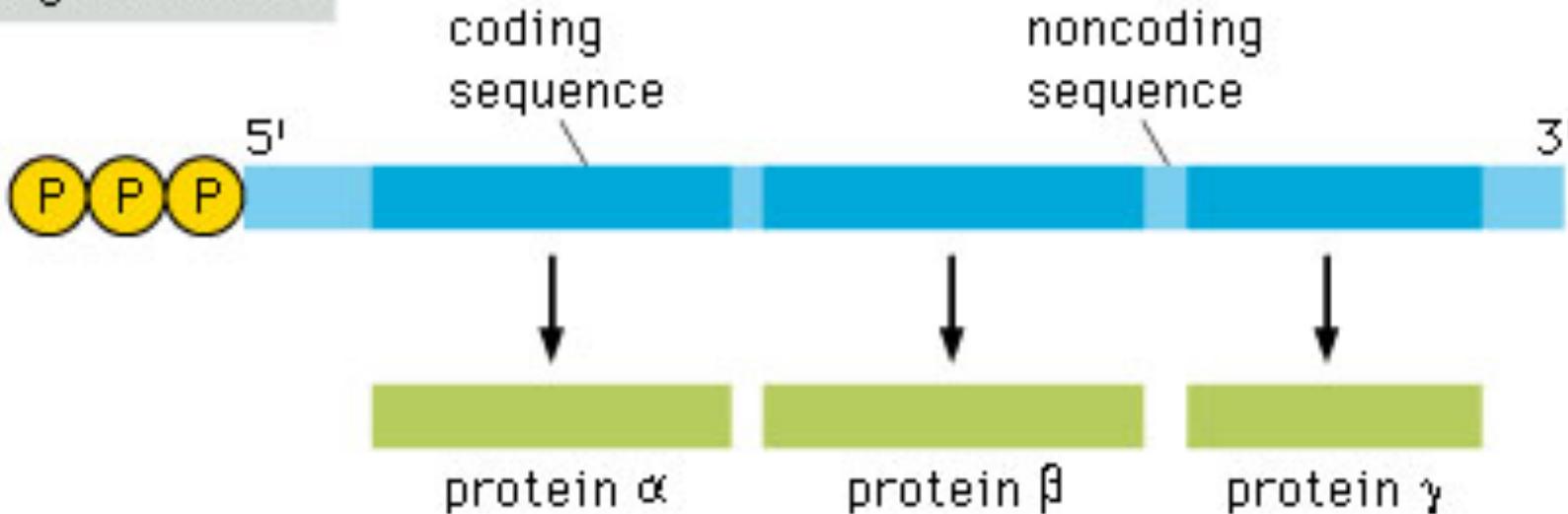


Eukaryotic Gene Expression

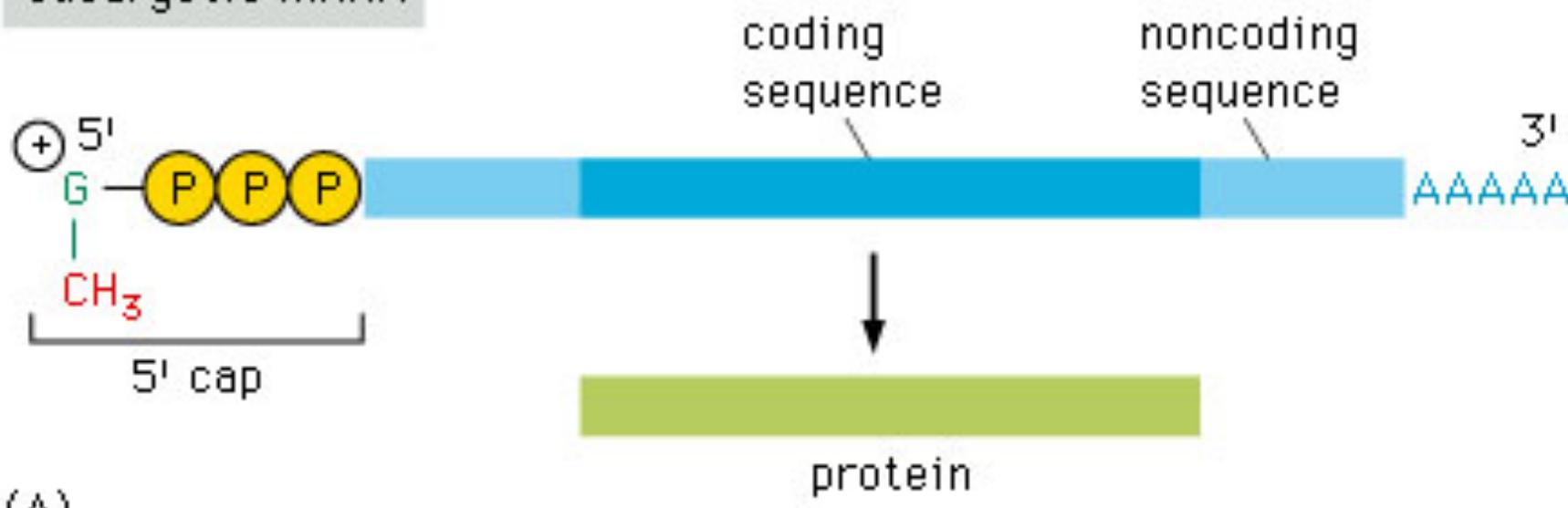
- The Promoter and the Transcription Machinery
- Transcriptional Enhancers and Repressors
- Epigenetic Control: Regulating Access to Genes within the Chromosome
- RNA Splicing
- The Initiation Complex and Translation Rate
- Regulating Protein Activity and Longevity



procaryotic mRNA



eucaryotic mRNA



(A)

Stages of Protein Synthesis in *E. coli*

- Is divided into three phases:
- Initiation
 - ribosome is assembled in the correct place on an mRNA ready to commence protein synthesis.
- Elongation
 - the correct amino acid is brought to the ribosome, is joined to the nascent polypeptide chain, and the entire assembly moves one position along the mRNA.
- Termination
 - happens when a stop codon is reached, there is no amino acid to be incorporated and the entire assembly dissociates to release the newly-synthesized polypeptide.
- There are two rules about protein synthesis to keep in mind:
 - mRNA is translated 5' -> 3'
 - Proteins are synthesized from the N-terminus to the C-terminus

The Process of Translation (protein synthesis) in both Prokaryotes and Eukaryotes

2nd base in codon

	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

1st base in codon

3rd base in codon

(1) ...AUAA**AGGAGG**UAAAUA**AUG** →

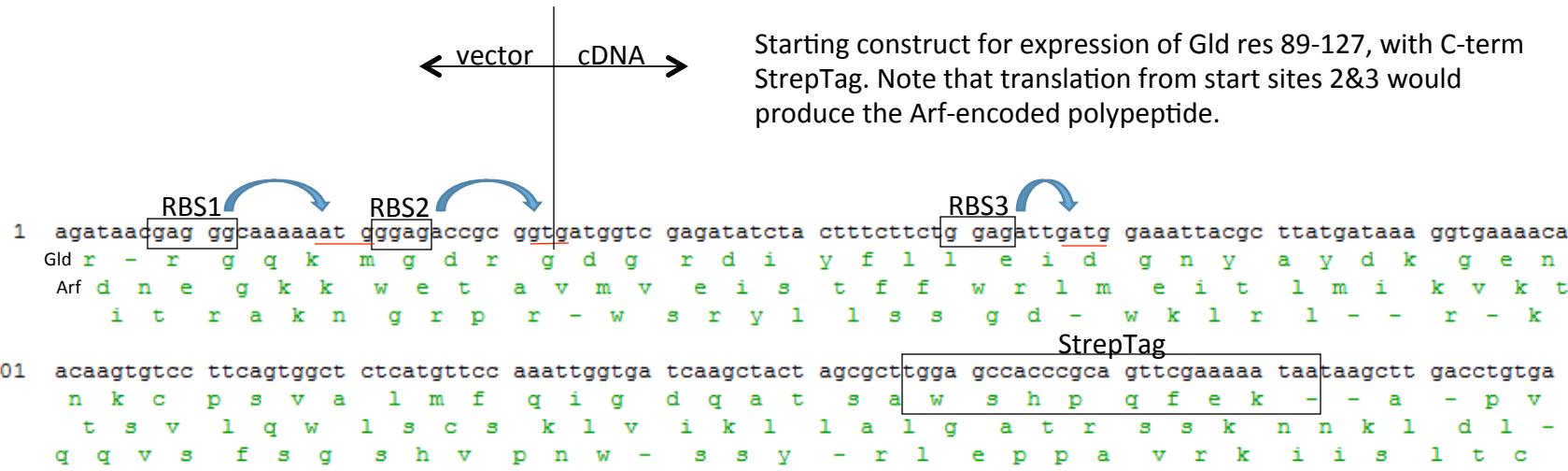
(2) .AUAA**AGGAAA**UAAAUA**AUG** →

.AUAA**ACA****GAGG**UAAAUA**AUG** →

.AUAA**ACA****GGAG**UAAAUA**AUG** →

Shine-Dalgarno start codon

The diagram illustrates the process of protein synthesis at a ribosome. A green ribosome is shown with its large and small subunits. A yellow mRNA strand is bound to the ribosome, with red arrows indicating its movement through the P site and A site. Blue tRNA molecules are shown bringing amino acids (represented by colored spheres) into the A site. A red line traces the growing polypeptide chain as it exits the ribosome. Labels include: newly born protein, amino acids, large subunit, tRNA, P site, A site, mRNA, and small subunit.



Arabidopsis Gld protein (GenBank NP_191796.1)

```

1 mlqlvvvawl liftaetvvp hqesgewsce sdseiqvlad frpgiitldg hnndwkdidg
61 sefplrpald pdsdheydag kmrvkalhdg rdiyflleid gnyaydkgen nkcpvalmf
121 qigdqatyhn mggckegtds ctskacrgfe vdimhfsign aipgrlyggn pidngegngg
181 drfghlvdiy awnphcryld glgpsgndss aqndwhgaww hssfttsgy veedspypd
241 gkkgtyyfef srplrtmdrl qqdvqftlgs pakmsvafwy pvdskpwhgs ghytincdwt
301 pldissgsss gltastvkgs sdgasitail lsmislvfsg fiayrlfspr nvpftpmmn
361 1

```

alternate reading frame (ARF) polypeptide

```

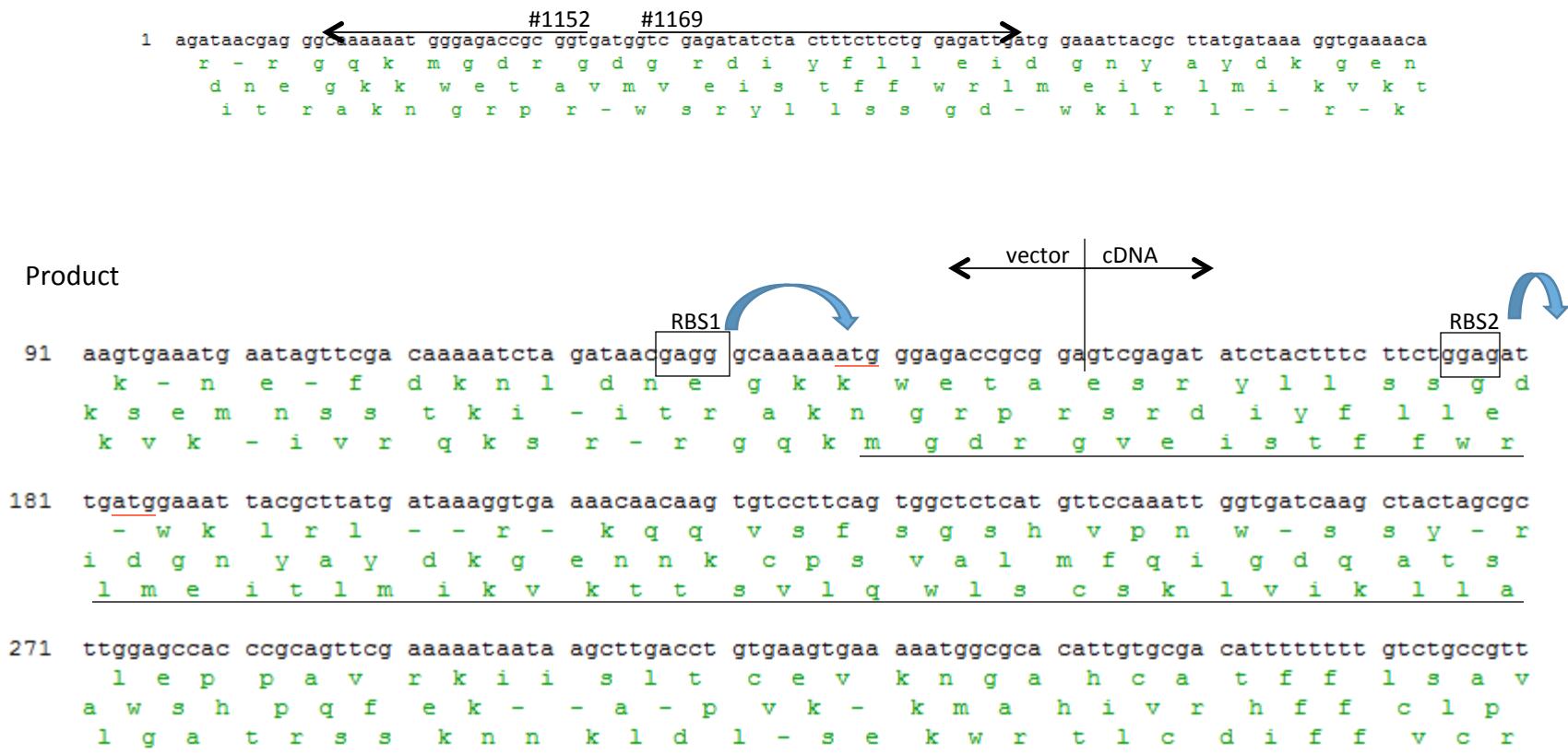
1 vmveistffw rlmeitlmik vkttsvlqlwl scsklvikll algatrsskn nkldl

```

↑
RBS3

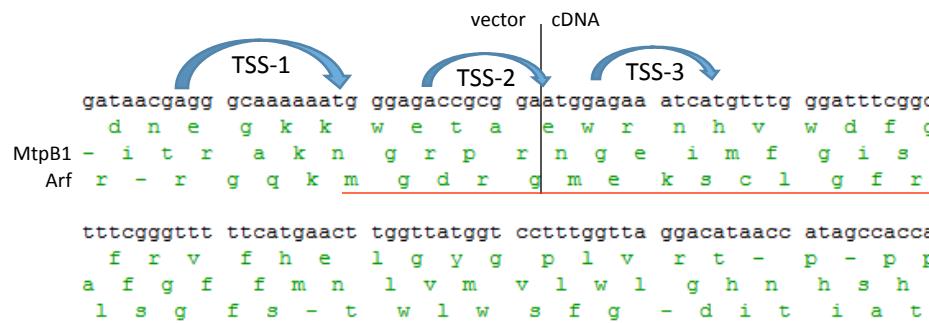
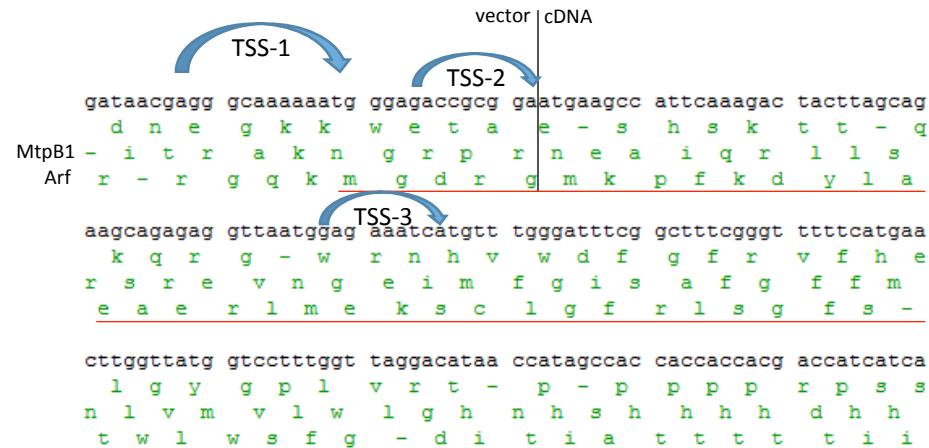
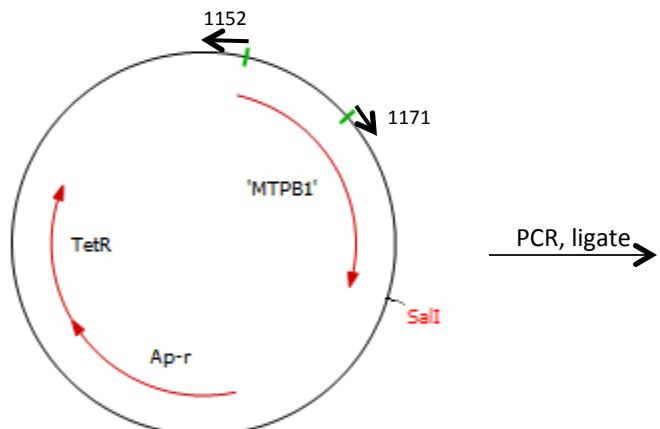
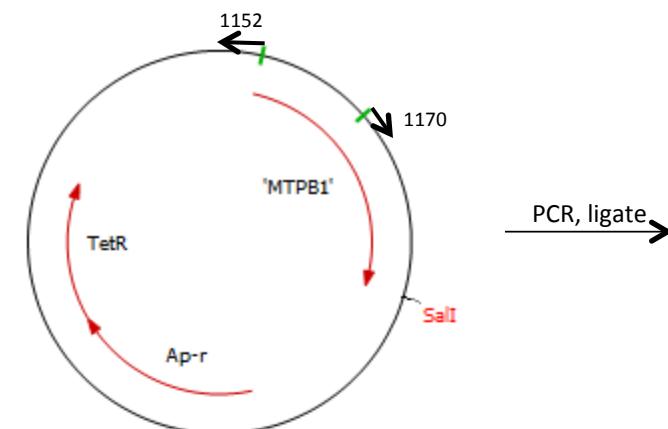
To confirm the frame 2-encoded polypeptide, but not Gld, is the toxic product, the following construct was made:

PCR



Results: expression of this product was as toxic as the starting construct shown in slide 1. Note there is no start site in this product for translation of the Gld polypeptide (now encoded in frame 2). Only the Arf polypeptide can be produced.

Toxic MtpB1 Arf polypeptides (earlier deletion analysis showed middle third of MtpB1 coding sequence was sufficient for toxic activity)



Top peptide (28mer) is strongly toxic, arresting growth immediately after induction. The truncated variant (bottom) lacking res 1-14 is less toxic, arresting growth after ~ 0.5 cell doublings. Note that both Arf peptides can be translated from start sites 1&2, whereas translation from TSS-3 would produce a truncated MtpB1 protein lacking first ~100 aa.

Cryptic translation start sites in the *Arabidopsis* MtpB1 cDNA (1-1197)

nt	SD	downstream sequence	TSS?	Frame
52	GAGG	AAGAA ATG GAGTCACCAT	Y	1
88	GAGG	AAAATTAGGC GTG GTTC	Y?	1
236	GAGG	TTTTAAGGCGAATAGTCT	N	-
478	GAGG	TTAATGGAGAAAT ATGT	Y	2
634	GAGG	TG GTG GCAGAACAGAGGAGG	?	1
649	GAGG	AGGAGGGAGGAA ATGA ATC	?	1
652	GAGG	AGGAGGGAA ATG AATCCTT	Y	1
655	GAGG	AGGAA ATG AATCCTT	Y	1
658	GAGG	AA ATG AATC	?	1
779	GAGG	TATCATTGGGTAAAACC	N	-
941	GAGG	ACTTAAGCGAATCG ATGG	?	2
1135	GAGG	TCGACCTGCAGGGGGACC	N	-

nt	SD	downstream sequence	TSS?	Frame	
4	GGAG	ACCGCGGAGAATTGGAGC	N	-	
13	GGAG	AATTGGAGCAAATCTGCA	N	-	
21	GGAG	CAAATCTGCATTTGAAA	N	-	
63	GGAG	TCACCATCACCATCGAAA	N	-	
235	GGAG	GTTTAAGGCGAATAGTC	N	-	
430	GGAG	TGATAATCC ATG AAGCCA	Y	2	Toxic
487	GGAG	AAAT CATG TTTGGGATT	Y	1	
633	GGAG	GTG GTG GCAGAAGAGGAC	Y	1	
651	GGAG	GAGGAGGAA ATG AATCCT	Y	1	
654	GGAG	GAGGAA ATG AATCCTTTA	Y	1	
657	GGAG	GAA ATG AATCCTTTAAAA	Y	1	
699	GGAG	ATG AACATTAACATACAA	?	1	
721	GGAG	CTTATCTTCACGCAATGG	N	-	
778	GGAG	GTATCATTGGGTAAAAC	N	-	
961	GGAG	TTAACGATTGTTT ATG ATC	?	2	
993	GGAG	ATAACAGTAGGGAGAAC	N	-	
1007	GGAG	AATCGTATTGTCCTGTCA	N	-	
1045	GGAG	CTAGTCCTAAAGAGATCA	N	-	
1172	GGAG	GGAGGGG A TTGGAAA	N	-	

Cryptic TSS in the *Arabidopsis* Gld cDNA (1-993)

404	GAGG	CTTCGAGGTTGAATA	N	-	
412	GAGG	TTGATATT ATG CATT	Y	1	
560	GAGG	TGACAGATTGGTCA	N	-	
658	GAGG	AAGACAGCCCTTACA	N	-	
726	GAGG	ACAATGGACC GTCTC	Y	1	
4	GGAG	ACCGCG GTG TTGTAC	Y	2	
44	GGAG	TT GTG AATCGCACTC	?	2	
63	GGAG	ATTCAGGTCTTAGCC	N	-	
255	GGAG	ATTG ATG GAAATTAC	Y	2	Toxic
499	GGAG	GTG ACAGATTGGTC	?	2	
613	GGAG	CGTG GTG GCACAGCA	Y	3	
968	GGAG	CCACCCGCAGTTCGA	N	-	

Cryptic TSS in the *E. coli* RpoB coding sequence (1-4029)

945	GGAG	CTGAGCCTGGATCTGCTGGCT	N	-	
2994	GGAG	CTGGGCCTGACAGACGAAG	N	-	
3813	GGAG	ATGGAA GTGTG GGC	Y	1, 3	
3945	GGAG	CCGGGC ATG	Y	1	

Summary & Conclusions

- Hypothesis: incompatibility between prokaryotic and eukaryotic gene expression
- Prokaryotic genes: No translation initiation start sites (TSS)
- Eukaryotic genes (have many TSS)
- A widen bottleneck in using gene expression and elimination of TSS.
- Future solution:
- To make synthetic genes by eliminating internal TSS and reducing codon bias

